

2018 吉林省大学生数学建模竞赛

承 诺 书

我们仔细阅读了《全国大学生数学建模竞赛章程》和《全国大学生数学建模竞赛参赛规则》（以下简称为“竞赛章程和参赛规则”，可从全国大学生数学建模竞赛网站下载）。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛章程和参赛规则的，如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛章程和参赛规则，以保证竞赛的公正、公平性。如有违反竞赛章程和参赛规则的行为，我们将受到严肃处理。

我们授权全国大学生数学建模竞赛组委会，可将我们的论文以任何形式进行公开展示（包括进行网上公示，在书籍、期刊和其他媒体进行正式或非正式发表等）。

我们参赛选择的题号是（从 A/B/C/D 中选择一项填写）：_____ B _____

我们的参赛报名号为（如果赛区设置报名号的话）：_____ 201807007000 _____

所属学校（请填写完整的全名）：_____ 东北电力大学 _____

参赛队员（打印并签名）：1. _____ 贾舒越 _____

2. _____ 杜科 _____

3. _____ 李雪 _____

指导教师或指导教师组负责人（打印并签名）：_____

（论文纸质版与电子版中的以上信息必须一致，只是电子版中无需签名。以上内容请仔细核对，提交后将不再允许做任何修改。如填写错误，论文可能被取消评奖资格。）

日期：_____ 2018 年 5 月 2 日 _____

赛区评阅编号（由赛区组委会评阅前进行编号）：

“低保标准”的数学模型建立

摘要

我国最低生活保障制度已经存在了近二十年，覆盖了全国大多数省市及地区，有越来越多的人受益。本文对 2017 年中国统计年鉴的数据进行主成分分析，筛选出影响“低保标准”的 6 项指标；基于 RBF 神经网络，建立了全国 31 个省市与对应的“低保标准”的预测模型，利用模型分析并确定了吉林省的“低保标准”；用主成分分析得出的“低保标准”的综合评价值相关性替代各地区“低保标准”相关性，通过阈值分类将各地区分为三类；基于地区分类，采用 RBF 神经网络建立同类高精度预测模型，并验证了模型适应性。

针对问题 1，对数据进行主成分分析，计算出各指标“低保标准”的贡献率，筛选出对“低保标准”累计影响高达 89.67% 的前 6 个主成分，第一主成分主要反映物价指数，第二个主成分主要反映人口老龄化指数，第三个指数主要反映人均生产总值，第四个主成分主要反映人均可支配收入，第五个主成分主要反映居住指数，最后一个主成分主要反映家庭户数指数。最后，我们得出对“低保标准”影响程度较大的 6 项指标：人均生产总值、人均可支配收入、居住指数、家庭户数指数、物价指数以及人口老龄化指数。

针对问题 2，通过整理出的 2016 年全国各省市城乡“低保标准”表，用问题 1 中的 6 项指标作为输入，30 个省市数据作为测试集，建立基于 RBF 神经网络的“低保标准”预测模型，最后采用吉林省与上海市的数据作为测试集测试，验证了模型的可靠性。

针对问题 3，用主成分分析得到“低保标准”的综合评价值 Z ，用其相关性替代各地区“低保标准”相关性。将 Z 的数据绘成图，得出“低保标准”相关性较大的省市，经过多次数据统计分析决定将阈值设置为 0.9999 和 -0.9999，评价结果在同一区间内的省市“低保标准”相关性高。根据阈值按照相关性将各省市分为三类：上海、北京、天津、江苏为 I 类；甘肃、西藏为 III 类；其他地区为 II 类。通过该模型给出的结果经过多方验证具有可靠性。

针对问题 4，根据问题 3 相关性的分类，对具有普遍代表性的 II 类建立新的 RBF 神经网络模型，而对 I 类、III 类仍采用已经建立的 RBF 神经网络模型，针对新的模型采用吉林省、西藏自治区的数据进行适应性验证。

关键词： 低保标准；主成分分析；神经网络；计算机模拟；阈值分类

1. 问题的重述

城市居民最低生活保障制度，是政府对城市贫困人口按最低生活保障标准进行差额救助的新型社会救济制度，是对我国传统社会救济制度的重大改革，它与退休职工基本养老保险制度、下岗职工基本生活保障和失业保险制度并称为“三条保障线”。这为中国社会主义市场经济的发展与社会稳定提供了必要的保障平台^[1]。从上世纪 90 年代的初创、试点、确立，再到如今全面法制化，最低生活保障制度已经存在了二十多年。“低保”制度覆盖了全国大多数省市及地区，在我国，不管是城市居民，还是农民居民，只要符合“低保”要求，都可以申请“低保”保障。我国“低保”制度在城市和农村并不相同，我国对农村低保制度的探索，实际上还早于城市。只不过受传统农村集体福利思维定式的束缚和农村税费改革的影响，此项制度建设一直进展缓慢。近年来国家发展快速，中央政府加大举措，农村“低保”制度也在趋于完善。截止到现在，全国城乡“低保”对象已经达到了 7487.4 万人。但是，由于各地区发展的不平衡等因素，每个省市甚至县级城镇的“低保”标准仍存在一定差异^{[1][2]}。

为实现精算扶贫、精准脱贫，有效提高贫困家庭的生活质量，我们将研究以下问题：

问题 1：通过搜集资料和查阅文献，列出影响“低保标准”的各种指标，筛选出计算“低保标准”的主要指标。

问题 2：根据问题 1 所选择指标，建立“低保标准”的数学模型。并根据模型，结合某省的数据，确定该省份的“低保标准”，进行分析说明。

问题 3：利用数据资料，分析出现行各地“低保标准”间的相关程度。

问题 4：利用问题 3 给出的分析结果，建立多元数学模型，最后进行模型适应性的验证。

2. 模型的准备

根据 2017 年中国统计年鉴提供的各省份（除港澳台外）的行政区划、国民经济和社会发展综合资料，我们整理并筛选出了与“低保标准”密切相关的 13 个因素：人均生产总值、居民消费水平指数、人均可支配收入、食品指数、衣着指数、居住指数、生活用品及服务指数、交通和通信指数、教育文化与娱乐指数、医疗保健指数、家庭户数指数、物价指数、人口老龄化指数^[2]。

(1) 指数的内容

表 1 影响“低保标准”的各项指标内容

	指标名称	指标内容
x_1	人均国内生产总值	一个国家核算期内（通常是一年）实现的国内生产总值与这个国家的常住人口（或户籍人口）比值。
x_2	居民消费水平指数	居民用于满足家庭日常生活消费需要的全部支出，既包括现金消费支出，也包括实物消费支出。

x_3	居民可支配收入	居民可用于最终消费支出和储蓄的总和，即居民可用于自由支配的收入。
x_4	食品指数	恩格尔系数(Engel's Coefficient) 食品支出总额占个人消费支出总额的比重。
x_5	衣着指数	居民穿着有关的支出，包括服装、服装材料、鞋类、其他衣类及配件、衣着相关加工服务的支出占总消费支出的比例。
x_6	居住指数	指与居住有关的支出占总消费支出的比例，包括房租、水、电、燃料、物业管理等方面的支出，也包括自有住房折算租金。
x_7	生活用品及服务指数	指家庭及个人的各类生活品及家庭服务支出占总消费支出的比例。 包括家具及室内装饰品、家用器具、家用纺织品、家庭日用杂品、个人用品和家庭服务。
x_8	交通和通信指数	指用于交通和通信工具及相关的各种服务费、维修费和车辆保险等支出占总消费支出的比例。
x_9	教育文化与娱乐指数	指用于教育、文化和娱乐方面的支出占总消费支出的比例。
x_{10}	医疗保健指数	指用于医疗和保健的药品、用品和服务的总费用占总消费支出的比例。包括医疗器具及药品，以及医疗服务。
x_{11}	家庭户数指数	家庭人口数量，具体可细分为1人户、2人户、3人户、4人户、5人以上户家庭。
x_{12}	物价指数	是反映一定时期内城乡商品零售价格变动趋势和程度的相对数。
x_{13}	人口老龄化指数	老年人口抚养比，也称老年人口抚养系数。

(2) 指数的计算公式

$$\text{食品指数}(R_1) = \frac{\text{食物支出变动}}{\text{总支出变动}} \times 100\%$$

$$\text{衣着指数} = \frac{\text{衣着消费支出}}{\text{消费总支出}} \times 100\%$$

$$\text{居住指数} = \frac{\text{用于居住消费支出}}{\text{消费总支出}} \times 100\%$$

$$\text{生活用品及服务指数} = \frac{\text{用于生活用品及服务消费支出}}{\text{消费总支出}} \times 100\%$$

$$\text{交通和通信指数} = \frac{\text{用于交通和通信工具消费支出}}{\text{消费总支出}} \times 100\%$$

$$\text{教育文化与娱乐指数} = \frac{\text{用于教育文化与娱乐消费支出}}{\text{消费总支出}} \times 100\%$$

$$\text{医疗保健指数} = \frac{\text{用于医疗保健消费支出}}{\text{消费总支出}} \times 100\%$$

$$\text{家庭户数指数}^{[2]} = \frac{\sum (\text{户数} \times r1)}{\text{分地区总户数}} \times 100\%$$

$$\text{人口老龄化指数: ODR} = \frac{P_{65+}}{P_{15\sim64}} \times 100\%$$

其中：ODR 为老年人口抚养比；

P_{65+} 为 65 岁及 65 岁以上的老年人口数；

$P_{15\sim64}$ 为 15~64 岁的劳动年龄人口数。

3. 问题的分析

为实现精算扶贫、精准脱贫、高效合理的帮助老百姓摆脱贫困，有效提高贫困家庭的生活质量，对于问题的详细分析如下：

针对问题 1：为了筛选出计算“低保标准”的主要指标，查阅相关文献与资料后，整理出了 13 项对“低保标准”造成影响的指标数据（附录一），进行前期准备与分析，推导并计算出影响指标的比例而不是单一的指标数据，利用主成分分析法，标准化处理原始数据，并计算相关系数矩阵，得出特征根与特征向量，计算出出各大主成分与各项指标的贡献率，筛选出主要影响“低保标准”的指标。

针对问题 2：问题 2 要求建立模型来给出某地区的“低保标准”。利用问题 1 基于主成分分析“低保标准”指标挑选模型，我们得到了贡献率最大的 6 项指标：人均生产总值、人均可支配收入、居住指数、家庭户数指数、物价指数以及人口老龄化指数。同时，经过大量查阅数据，筛选与整理出 2016 年全国各省市城乡低保标准表（附录二）。对 6 项核心指标影响下对应的城乡低保标准，用已知的 6 项指标作为输入，30 个省市数据作为测试集，建立基于 RBF 神经网络的“低保标准”预测模型，最后用吉林省的数据作为测试集测试模型的可靠性。

针对问题 3：问题 3 要求根据获得的数据，分析各省市之间“低保标准”的相关性。根据问题 1 得出的特征向量，计算出“低保标准”的综合评价值 Z，它可以直接反映出各省市“低保标准”的高低，在分析与比较每个值之间的大小后进行排序，利用阈值分类呈现出各省市相关性。

针对问题 4：问题 4 需要给出多元数学模型，并对模型进行适应性验证。根据问题 3 中的分类，按照特殊与一般的思想，其中特殊条件比如说国家政策、地理位置优越性等下的两类，例如北京、上海等“低保标准”遥遥领先的地区与甘肃、西藏“低保标准”远远落后的地区归结为一类，免去对问题二中建立的神经网络模型的可能干扰，得出多元强健的数学模型。同时我们利用吉林与西藏的 6 项数据进行预测，比较预测值与实际值之间的差值，评价模型的准确性。

4. 模型假设

- (1) 忽略掉在各省市占据很少比例的领取低保的残疾人或重症患者。
- (2) 筛选的数据中各省市的“低保标准”为其各地区“低保标准”的平均值。
- (3) 排除重大自然灾害，例如：特大地震、海啸等给人们带来的重大损失。

5. 符号说明

符号	含义
n	输入层节点数
p	隐含层节点数
q	输出层节点数
w_{ik}	隐含层第 i 个神经元与输出层第 k 个神经元的连接权值
$R_i(X)$	隐含层第 i 个神经元的作用函数
X	n 维输入向量
C_i	第 i 个基函数的中心，与 X 具有相同维数的向量
σ_i	第 i 个基函数的宽度
$\ X - C_i\ $	输入向量与阈值向量之间的距离
min_k	训练集中第 k 个输出神经元中所有期望输出的最小值
max_k	训练集中第 k 个输出神经元中所有期望输出的最大值。
$mini$	训练集中第 i 个特征所有输入信息的最小值
$maxi$	训练集中第 i 个特征所有输入信息的最大值。
$W_{kj}(t)$	第 k 个输出神经元与第 j 个隐含层神经元之间在第 t 次迭代计算时的调节权重
$c_{ji}(t)$	第 j 个隐含层神经元对于第 i 个输入神经元在第 t 次迭代计算时的中心分量
$d_{ji}(t)$	与中心 $c_{ji}(t)$ 对应的宽度
η	学习因子
E	RBF 神经网络评价函数
O_{lk}	第 k 个输出神经元在第 1 个输入样本时的期望输出值
y_{lk}	第 k 个输出神经元在第 1 个输入样本时的网络输出值

6. 模型的建立与求解

6.1 问题 1 的求解——基于主成分分析的“低保标准”指标挑选

主成分分析法是利用“降维”的思想，把多个指标转化为少数几个综合指标（即主成分），其中每个主成分都能够反映原始变量的大部分信息，且所含信息互不重复^[3]，我们通过主成分分析得出了 6 项对“低保标准”最具有代表性的指标。

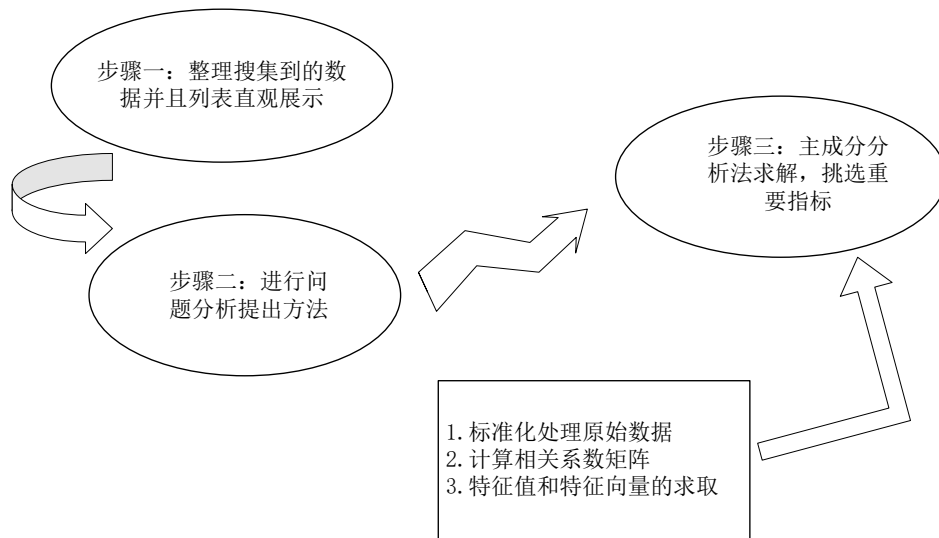


图 1 模型的求解思路

6.1.1 对原始数据进行标准化处理

假设进行主成分分析的指标变量有 m 个，分别为 x_1, x_2, \dots, x_m ，共有 n 个评价对象，第 i 个评价对象的第 j 个指标的取值为 a_{ij} 。将各个指标值 a_{ij} 转化成标准化指标值 \bar{a}_{ij} ，有

$$\bar{a}_{ij} = \frac{a_{ij} - \mu_j}{s_j} \quad (i=1, 2, \dots, n; \quad j=1, 2, \dots, m)$$

其中， $\mu_j = \frac{1}{n} \sum_{i=1}^n a_{ij}$ ， $s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (a_{ij} - \mu_j)^2}$ ($j=1, 2, \dots, m$)。

即 μ_j, s_j 为第 j 个指标的样本均值和样本标准差。对应地，称

$$\bar{x}_j = \frac{x_j - \mu_j}{s_j} \quad (j=1, 2, \dots, m)$$

为标准化指标变量。

6.1.2 计算相关系数矩阵 R

相关系数矩阵 $R = (r_{ij})_{m \times m}$ ，有

$$r_{ij} = \frac{\sum_{k=1}^n \bar{a}_{ki} \times \bar{a}_{kj}}{n-1} \quad (i=1, 2, \dots, m; \quad j=1, 2, \dots, m)$$

其中： $r_{ii}=1$ ， $r_{ij}=r_{ji}$ ， r_{ij} 是第 i 个指标与第 j 个指标的相关系数。

6.1.3 计算特征值和特征向量

计算相关系数矩阵 R 的特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$ 及对应的特征向量 u_1, u_2, \dots, u_m ，其中 $u_j = [u_{1j}, u_{2j}, \dots, u_{mj}]^T$ ，由特征向量组成 m 个新的指标变量：

$$y_1 = u_{11}\bar{x}_1 + u_{21}\bar{x}_2 + \dots + u_{m1}\bar{x}_m$$

$$y_2 = u_{12}\bar{x}_1 + u_{22}\bar{x}_2 + \dots + u_{m2}\bar{x}_m$$

$$y_m = u_{1m}\bar{x}_1 + u_{2m}\bar{x}_2 + \dots + u_{mm}\bar{x}_m$$

其中： y_1 是第 1 主成分， y_2 是第 2 主成分， \dots ， y_m 是第 m 主成分。

6.1.4 指标选择结果及分析

我们利用 Matlab 软件对 13 个评价指标进行主成分分析，相关系数矩阵的前 6 个特征根及其贡献率如下：

表 2 主成分分析结果

序号	特征根	贡献率	累计贡献率	序号	特征根	贡献率	累计贡献率
1	5.139449	39.53423	39.534227	4	1.165663	8.96664	77.4137114
2	2.234616	17.18935	56.723579	5	0.948162	7.293554	84.7072658
3	1.524054	11.72349	68.447071	6	0.645203	4.963102	89.6703678

可以看出，前两个特征根累计贡献率超过 50%，前 3 个特征根的累计贡献率就达到近 70%，主成分分析效果很好。

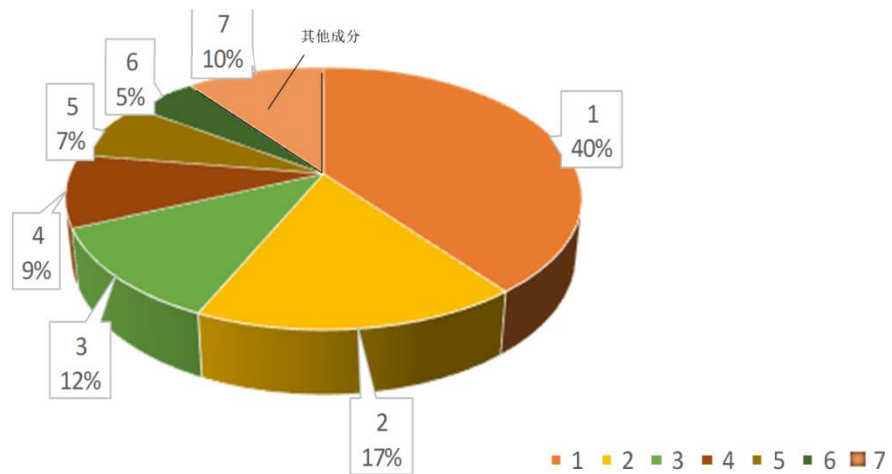


图 2 主成分贡献率比重

下面选取前 6 个主成分（累计贡献率近 90%）进行综合评价。

前 6 个特征根对应的特征向量如下：

表 3 标准化变量的前 6 个主成分对应的特征向量

序号	X1	X2	X3	X4	X5	X6	y
1	118198	52530.4	1.0226824	1.0146736	101.4	0.1174167	710
2	115053	34074.5	1.0146915	1.0097322	102.1	0.113598	705
3	43062	19725.4	0.9921182	0.9807925	101.5	0.1090837	500
4	35532	19048.9	0.9881306	0.9942916	101.1	0.0869621	479
5	72064	24126.6	0.9881423	1.0090344	101.2	0.0943887	540
6	50791	26039.7	0.9891732	1.0093293	101.6	0.1322406	542.5
7	53868	19967	0.9812992	1.0028842	101.6	0.1085204	392.5
8	40432	19838.5	0.9852217	1.0067435	101.5	0.1193763	510
9	116562	54305.3	1.0184109	1.0232611	103.2	0.1296315	790
10	96887	32070.1	0.9892473	0.9864593	102.3	0.1350488	700
11	84916	38529	0.9911678	1.0139434	101.9	0.1163546	660
12	39561	19998.1	0.9931238	0.9750489	101.8	0.114665	510
13	74707	27607.9	0.9901672	0.9904064	101.7	0.099495	700
14	40400	20109.6	0.9901961	0.9506182	102	0.0959664	465
15	68733	24685.3	0.9882468	1.0051857	102.1	0.1165173	525
16	42575	18443.1	1.0029441	0.9696466	101.9	0.1007612	520
17	55665	21786.6	1.0058708	0.9912611	102.2	0.115746	580
18	46382	21114.8	0.9931305	0.9810905	101.9	0.1188972	450
19	74016	30295.8	0.9941349	0.974316	102.3	0.0769323	600
20	38027	18305.1	0.9872047	0.9475188	101.6	0.0969344	250
21	44347	20653.4	0.9980545	0.9367574	102.8	0.082489	450
22	58502	22034.1	0.9931238	1.0100777	101.8	0.1396714	375
23	40003	18808.3	0.9931305	0.9958094	101.9	0.1369624	450
24	33246	15121.1	0.9940828	0.9706767	101.4	0.0962393	477.5
25	31093	16719.9	0.9970443	0.9628986	101.5	0.0837548	502.5
26	35184	13639.2	0.9834146	0.8684889	102.5	0.0498387	640
27	51015	18873.7	0.9960513	0.9871346	101.3	0.1074218	495
28	27643	14670.3	0.9950642	0.962879	101.3	0.0995446	451
29	43531	17301.8	1.0333988	0.9661251	101.8	0.0721877	373
30	47194	18832.3	0.9891626	0.9842286	101.5	0.0778327	380
31	40564	18354.7	0.9980276	0.9718639	101.4	0.0723035	380

由此可得 6 个主成分分别为：

$$y_1 = 0.406047142x_1 + 0.149280486x_2 + \dots + -0.27716208x_{13}$$

$$y_2 = 0.414015413x_1 + 0.11855429x_2 + \dots + 0.716542347x_{13}$$

$$y_3 = 0.40470842x_1 + 0.100723003x_2 + \dots + -0.500612192x_{13}$$

$$y_4 = -0.05344379x_1 + -0.27614783x_2 + \dots + -0.15951067x_{13}$$

$$y_5 = -0.19021636x_1 + 0.365169398x_2 + \dots + -0.019972731x_{13}$$

$$y_6 = 0.278708514x_1 + -0.17946803x_2 + \dots + -0.260303399x_{13}$$

最后，我们得出了对“低保标准”影响程度较大的 6 项指标：人均生产总值、人均可支配收入、居住指数、家庭户数指数、物价指数以及人口老龄化指数。把各省市原始 13 个指标的标准化数据代入 6 个主成分的表达式，就可以得到各省市的 6 个主成分值，这一结果将为解决题目中第三个问题建立综合评价模型埋下伏笔。

(1) 人均国内生产总值与“低保标准”息息相关。我们在查阅相关数据后发现，低保资金与人均国内生产总值成正相关，人均国内生产总值越高，低保资金越多；人均国内生产总值越低，低保资金越少。在当今时代中国经济迅猛发展的大背景下，人均国内生产总值（人均 GDP）也在不同程度增长，我国的“低保标准”也在不同程度上调^[1]。

(2) 居民可支配收入也在一定程度影响着“低保标准”。我们在查阅相关数据后发现，低保的对象中有 70% 以上是失业人员，可支配收入寥寥无几。失业使他们几十年间形成的生活方式和生产方式都发生了巨变，其中影响最大的就是“不稳定”。低保制度使他们重新有了一笔月月可领取的“稳定收入”，贫困家庭就不会“心慌”。所以，低保制度所起到的作用，在社会心理方面比在物质方面更大。居民可支配收入在一定程度变化，“低保标准”也随之一定程度变化。

(3) 居住指数越低，居民获得“低保标准”的可能性越大，“低保标准”也越高。当今房价居高不下，出现了“买房难、住房难”的一系列问题，“低保标准”也为很多人减少住房或还钱的压力。

(4) 家庭户数指数是家庭人口数量，我们通过查阅相关数据，根据国家统计局城市调查总队 1997 年数据，1 人户、2 人户、3 人户、4 人户、5 人以上户家庭的规模影响系数 r_1 分别是 1.13、1.01、1、0.98 和 0.94^[2]。

(5) 物价指数是反映一定时期内城乡商品零售价格变动趋势和程度的相对数。商品零售价格的变动与国家的财政收入、市场供需的平衡、消费与积累的比例关系有关。因此，该指数可以从一个侧面对上述经济活动进行观察和分析。物价是影响低保待遇变动的核心，抛开家庭情况、地区经济发展和财政收支因素，物价变动带来的低保待遇购买力变动程度是调整比例的底线，其余要素都在这个基础上进行调整。因此低保待遇比例调整的实质就是为了弥补因物价上涨带来的低保人群购买力损失^[2]。

(6) 我国人口老龄化问题严峻，如何解决一大批无经济来源的老人生活问题，“低保标准”显得尤为关键。人口老龄化指数也在我们的主成分分析中对“低保标准”占据很大的贡献率。

我们得出的结论也正与各个省市“低保标准”的实际情况相符合，进一步验

证了我们的主成分分析模型的具有的优点：

1. 少数重要指标替代大多数指标

影响“低保标准”的成分众多，我们精心挑选的 13 项指标作为“低保标准”的衡量因子。我们在主成分分析对原指标变量进行变换后将 13 项影响“低保标准”的指标形成了彼此相互独立的主成分，保留绝大部分信息的情况下用少数几个综合指标代替原指标进行分析，而且实践证明指标之间相关程度越高，主成分分析效果越好^[3]。

2. 普遍适用性

在对最新的国家低保政策以及 2016 年最新各地城乡进行了查阅和研究后，我们发现政策在趋向于城乡一体化与区域协调发展，农村的“低保标准”逐渐趋近于城市的“低保标准”，所以我们选取了最新的 2016 年各项指标数据进行分析，同时，我们选取了中国各地区 31 个省市的数据，能够对全国的普遍“低保标准”进行有效分析与评价。

这种方法在引进多方面变量的同时将复杂因素归结为几个主成分，使问题简单化，同时得到的结果更加科学有效的数据信息。在实际问题研究中，为了全面、系统地分析问题，我们必须考虑众多影响因素。这些涉及的因素一般称为指标，在多元统计分析中也称为变量。因为每个变量都在不同程度上反映了所研究问题的某些信息，并且指标之间彼此有一定的相关性，因而所得的统计数据反映的信息在一定程度上有重叠。主要方法有特征值分解，SVD，NMF 等^[4]。

6. 2 问题 2 的求解——基于 RBF 神经网络的“低保标准”预测模型

我们已经由主成分分析得出主要影响“低保标准”的 6 项指标：人均生产总值、人均可支配收入、居住指数、家庭户数指数、物价指数以及人口老龄化指数。接下来，我们基于 RBF 神经网络，利用全国各省市利用 6 项指标作为输入与对应的“低保标准”来建立一个强健的“低保标准”预测模型。

6. 2. 1 RBF 神经网络结构

RBF 是具有单隐层的三层前向网络。

第一层为输入层，由信号源节点组成。

第二层为隐藏层，隐藏层节点数视所描述问题的需要而定，隐藏层中神经元的变换函数即径向基函数是对中心点径向对称且衰减的非负线性函数，该函数是局部响应函数，具体的局部响应体现在其可见层到隐藏层的变换跟其它的神经网络不同。以前的前向网络变换函数都是全局响应的函数。

第三层为输出层，是对输入模式做出的响应。

输入层仅仅起到传输信号作用，输入层和隐含层之间之间可以看做连接权值为 1 的连接，输出层与隐含层所完成的任务是不同的，因而他们的学习策略也不同。输出层是对线性权进行调整，采用的是线性优化策略，因而学习速度较快；而隐含层是对激活函数（一般为高斯函数）的参数进行调整，采用的是非线性优化策略，因而学习速度较慢。

RBF 神经网络有很强的逼近能力、分类能力和学习速度。其工作原理是把网络看成对未知函数的逼近，任何函数都可以表示成为一组基函数的加权和，也即选择各隐层神经元的传输函数，使之构成一组基函数来逼近未知函数。

设输入层的输入为 $X=[x_1, x_2, \dots, x_n]$ ，实际输出为 $Y=[y_1, y_2, \dots, y_q]$ 。输入层实现从 $X \rightarrow R_i(X)$ 的非线性映射，输出层实现 $R_i(X) \rightarrow y_k$ 的线性映射，输出层第 k

个神经网络输出为

$$\hat{y}_k = \sum_{i=1}^p w_{ik} R_i(X) \quad (k=1, \dots, q)$$

$R_i(X)$ 表达式为:

$$R_i(X) = \exp(-\|X - C_i\|^2 / 2\sigma_i^2) \quad (i=1, \dots, p)$$

$R_i(X)$ 在 C_i 处有唯一的最大值, 随着 $\|X - C_i\|$ 的增大, $R_i(X)$ 迅速衰减到 0。对于给定的输入, 只有一小部分靠近 X 的中心被激活。当确定了 RBF 网络的聚类中心 C_i 、权值 w_{ik} 以后, 就可以求出给定某一输入时, 网络对应的输出值。

6.2.2 网络学习算法

RBF 神经网络学习算法需要三个参数: 基函数的中心 C_i , 方差(宽度) σ_i 以及隐含层到输出层的权值 w_{ik} 。在 RBE 网络中, 隐层执行的是一种固定不变的非线性变换, C_i, σ_i, w_{ik} 需通过学习和训练来确定。我们下面对算法进行分析。

1. RBF 神经网络中心选取方法:

对于 RBF 神经网络的学习算法, 关键问题是隐藏层神经元中心参数的合理确定。可以采用自组织学习选取 RBF 中心, RBF 神经网络的中心可以变化, 通过自组织学习确定其位置。输出层的线性权重则是通过有监督的学习来确定的。这种方法是对神经网络资源的再分配, 通过 学习, 使 RBF 的隐含层神经元中心位于输入空间重要的区域。这种方法主要采用 K-均值聚类法来选择 RBF 的中心, 属于无监督(导师)的学习方法。

2. 基于高斯核的 RBF 神经网络拓扑结构:

第一层输入层: 由信号源节点构成, 仅起到数据信息的传递作用, 对输入信息不做任何变换

第二层隐含层: 节点数视需要而定。隐含层神经元核函数(作用函数)是高斯函数, 对输入信息进行空间映射的变换。

第三层输出层, 对输入模式做出响应。输出层神经元的作用函数为线性函数, 对隐含层神经元输出的信息进行线性加权后输出, 作为整个神经网络的输出结果。

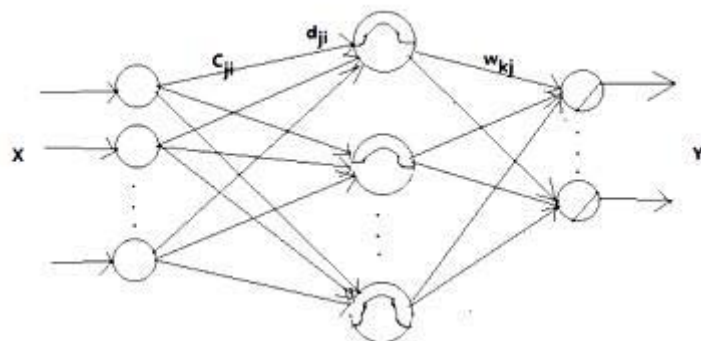


Fig. 3 基于高斯核的 RBF 神经网络的拓扑结构

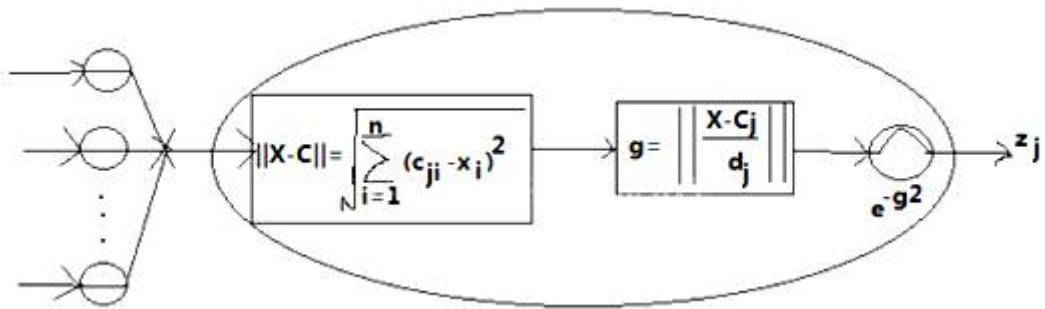


Fig. 4 径向基神经元模型机构

径向基神经网络传递参数采用 Gaussian 函数的方式：

$$\varphi_i(t) = e^{-\frac{t^2}{\delta_i^2}} \quad \text{即} \quad y = e^{-x^2}$$

当输入自变量为 0 时，传递函数取得最大值 1。随着权值和输入向量间的距离不断减小，网络输出是递增的。也就是说，径向基函数对输入信号在局部产生响应。函数的输入信号 X 靠近函数的中央范围时，隐含层节点将产生较大的输出。由此可以看出这种网络具有局部逼近能力。

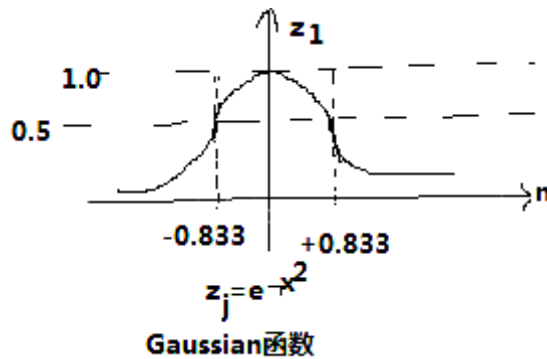


Fig. 5 高斯函数

当输入向量加到网络输入端时，径向基层每个神经元都会输出一个值，代表输入向量与神经元权值向量之间的接近程度。如果输入向量关于权值向量相差很多，则径向基层输出接近于 0；如果输入向量与权值向量很接近，则径向基层的输出接近于 1，经过第二层（隐含层）的线性神经元，输出值就靠近第二层权值。在这个过程中，如果只有一个径向基神经元的输出为 1，而其他神经元输出均为 0 或者接近 0，那么线性神经元的输出就相当于输出为 1 的神经元对应的第二层（隐含层）权值的值。

3. RBF 网络训练：

训练的目的在于求两层的最终权值 C_j 、 D_j 和 W_j 。

训练的过程分为两步：第一步是无监督学习，训练确定输入层与隐含层间的权值 C_j 、 D_j ；第二步是有监督学习，训练确定隐含层与输出层间的权值 W_j 。

训练前提供输入向量 X、对应的目标输出向量 Y 和径向基函数的宽度向量 D_j 。在第 1 次输入样品 ($i=1, 2, \dots, N$) 进行训练时，各个参数的表达及计算方法如下：

(1) 确定参数

①确定输入向量X:

$$X = [x_1, x_2, \dots, x_n]^T$$

②确定输出向量Y和希望输出向量 O

$$Y=[y_1, y_2, \dots, y_q]^T$$

$$O=[o_1, o_2, \dots, o_q]^T$$

③初始化隐含层至输出层的连接权值

$$O=[w_{k1}, w_{k2}, \dots, w_{kp}]^T, (k=1, 2, \dots, q)$$

参考中心初始化的方法给出隐藏层到输出层的权值初始化方法:

$$W_{kj}=\text{mink}+j\frac{\text{maxk}-\text{mink}}{q+1}$$

④初始化隐含层各神经元的中心参数 $C_j=[c_{j1}, c_{j2}, \dots, c_{jn}]^T$ 。

基于上述四项, RBF 神经网络中心参数的初始值为:

$$C_{ji}=\text{mini} + \frac{\text{maxi}-\text{mini}}{2p} + (j-1) \frac{\text{maxi}-\text{mini}}{p}$$

(p 为隐含层神经元总个数, $j=1, 2, \dots, p$)

⑤初始化宽度向量 $D_j = [d_{j1}, d_{j2}, \dots, d_{jn}]$ 。宽度向量影响着神经元对输入信息的作用范围: 宽度越小, 相应隐含层神经元作用函数的形状越窄, 那么处于其他神经元中心附近的信息在该神经元出的响应就越小。计算方法:

$$d_{ji}=d_f \sqrt{\frac{1}{N} \sum_{k=1}^N (x_i^k - C_{ji})^2}$$

d_f 为宽度调节系数, 取值小于 1, 作用是使每个隐含层神经元更容易实现对局部信息的感受能力, 有利于提高 RBF 神经网络的局部响应能力。

(2) 计算隐含层第 j 个神经元的输出值 z_j

$$z_j = -\exp\left(-\left\|\frac{x-C_j}{D_j}\right\|\right) \quad (j=1, 2, \dots, p)$$

C_j 是隐含层第 j 个神经元的中心向量, 由隐含层第 j 个神经元对应于输入层所有神经元的中心分量构成, $C_j=[c_{j1}, c_{j2}, \dots, c_{jn}]^T$; D_j 为隐含层第 j 个神经元的宽度向量, 与 C_j 相对应, $D_j = [d_{j1}, d_{j2}, \dots, d_{jn}]$, D_j 越大, 隐含层对输入向量的影响范围就越大, 且神经元间的平滑度也比较好; $\|\cdot\|$ 为欧式范数。

(3) 计算输出层神经元的输出

$$Y=[y_1, y_2, \dots, y_q]^T$$

$$y_k = \sum_{j=1}^p w_{kj} z_j \quad (k=1, 2, \dots, q)$$

其中: w_{kj} 为输出层第 k 个神经元与隐含层第 j 个神经元间的调节权重。

(4) 权重参数的迭代计算

RBF 神经网络权重参数的训练方法在这里取为梯度下降法。中心、宽度和调节权重参数均通过学习来自适应调节到最佳值, 迭代计算如下:

$$W_{kj}(t) = w_{kj}(t-1) + \eta \frac{\partial E}{\partial w_{kj}(t-1)} + \alpha [w_{kj}(t-1) - w_{kj}(t-2)]$$

$$c_{ji}(t) = c_{ji}(t-1) + \eta \frac{\partial E}{\partial c_{ji}(t-1)} + \alpha [c_{ji}(t-1) - c_{ji}(t-2)]$$

$$d_{ji}(t) = d_{ji}(t-1) + \eta \frac{\partial E}{\partial d_{ji}(t-1)} + \alpha [d_{ji}(t-1) - d_{ji}(t-2)]$$

E 为 RBF 神经网络评价函数：

$$E = \frac{1}{2} \sum_{l=1}^N \sum_{k=1}^q (y_{lk} - O_{lk})^2$$

综上所述，给出 RBF 神经网络的学习算法：

① 按(1)确定参数的五个步骤对神经网络参数进行初始化，并给定 η 和 α 的取值及迭代终止精度 ε 的值。

②按下式计算网络输出的均方根误差 RMS 的值，若 $RMS \leq \varepsilon$ ，则训练结束，否则转到第③步

$$RMS = \sqrt{\frac{\sum_{l=1}^N \sum_{k=1}^q (O_{lk} - y_{lk})^2}{qN}}$$

③按照(4)权重迭代计算，对调节权重，中心和宽度参数进行迭代计算。

④返回步骤②

6.2.3 网络学习样本的建立

经过大量查阅数据库与研究，通过《2016 年最新各地城乡低保标准表》，我们筛选与整理出 2016 年全国各省市城乡低保标准表，然后我们通过 6 项核心指标的影响下对应的城乡低保标准，我们用已知的 6 项指标作为输入，30 个省市数据作为测试集，建立基于 RBF 神经网络的“低保标准”预测模型，构成 6 个输入 1 个输出的网络，将前 30 个各省市 6 项核心指标数据作为训练样本集，后一个样本(吉林)的 6 项指标与“低保标准”数据作为预测检验样本。

表 4 神经网络学习样本

序号	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	y
1	118198	52530.4	1.0226824	1.0146736	101.4	0.1174167	710
2	115053	34074.5	1.0146915	1.0097322	102.1	0.113598	705
3	43062	19725.4	0.9921182	0.9807925	101.5	0.1090837	500
4	35532	19048.9	0.9881306	0.9942916	101.1	0.0869621	479
5	72064	24126.6	0.9881423	1.0090344	101.2	0.0943887	540
6	50791	26039.7	0.9891732	1.0093293	101.6	0.1322406	542.5
7	53868	19967	0.9812992	1.0028842	101.6	0.1085204	392.5
8	40432	19838.5	0.9852217	1.0067435	101.5	0.1193763	510
9	116562	54305.3	1.0184109	1.0232611	103.2	0.1296315	790
10	96887	32070.1	0.9892473	0.9864593	102.3	0.1350488	700
11	84916	38529	0.9911678	1.0139434	101.9	0.1163546	660

12	39561	19998.1	0.9931238	0.9750489	101.8	0.114665	510
13	74707	27607.9	0.9901672	0.9904064	101.7	0.099495	700
14	40400	20109.6	0.9901961	0.9506182	102	0.0959664	465
15	68733	24685.3	0.9882468	1.0051857	102.1	0.1165173	525
16	42575	18443.1	1.0029441	0.9696466	101.9	0.1007612	520
17	55665	21786.6	1.0058708	0.9912611	102.2	0.115746	580
18	46382	21114.8	0.9931305	0.9810905	101.9	0.1188972	450
19	74016	30295.8	0.9941349	0.974316	102.3	0.0769323	600
20	38027	18305.1	0.9872047	0.9475188	101.6	0.0969344	250
21	44347	20653.4	0.9980545	0.9367574	102.8	0.082489	450
22	58502	22034.1	0.9931238	1.0100777	101.8	0.1396714	375
23	40003	18808.3	0.9931305	0.9958094	101.9	0.1369624	450
24	33246	15121.1	0.9940828	0.9706767	101.4	0.0962393	477.5
25	31093	16719.9	0.9970443	0.9628986	101.5	0.0837548	502.5
26	35184	13639.2	0.9834146	0.8684889	102.5	0.0498387	640
27	51015	18873.7	0.9960513	0.9871346	101.3	0.1074218	495
28	27643	14670.3	0.9950642	0.962879	101.3	0.0995446	451
29	43531	17301.8	1.0333988	0.9661251	101.8	0.0721877	373
30	47194	18832.3	0.9891626	0.9842286	101.5	0.0778327	380
31	40564	18354.7	0.9980276	0.9718639	101.4	0.0723035	380

6.2.4 原始数据的预处理

采用以下式子分别对样本的输入、输出数据进行标准化处理，有

$$\tilde{t} = \frac{2(t-t_{min})}{t_{max}-t_{min}} - 1$$

其中： t 为格式化前的变量； t_{max} 、 t_{min} 分别为 t 的最大值和最小值； \tilde{t} 为格式化后的变量。

利用 Matlab 中提供了对数据进行归一化处理的函数：

$$T = \text{mapminmax}('reverse', tn, ps)$$

执行的算法是

$$t=0.5(\tilde{t}+1) \times (t_{max} - t_{min}) + t_{min}$$

6.2.5 网络的训练与预测

Matlab 代码位于附录三。

由于“低保标准”中自变量有 6 个，因变量有 1 个，输入神经元的个数取为 4，输出神经元的个数取为 1，中间隐含层神经元的个数，BP 网络需要根据经验取定，RBF 网络会在训练过程中自适应的取定^[8]。

6.2.6 RBF 模型、计算结果以及解释

我们选取用于测试的是吉林省，理由如下：

1. 普遍性

吉林的各项指标以及我们在第三问建立的综合评价模型均处于全国各省市的中等位置，既避免了北京、上海、天津、江苏第一梯队作为全国改革开放重点发展城市的特殊性，也避免了西藏、云南、甘肃、新疆等省市第三梯队由于地理位置、科技发展等方面因素的局限性，作为中间梯队的吉林，恰好可以代表大部分省市的实际状况^[9]。

2. 与假设较为符合

经过查阅相关政府网站与资料，我们发现吉林省在“低保标准”政策方面以年为更换“低保标准”的节点，不会在季度或者月份改变标准；同时，经过调查相关数据，吉林省残疾人总人数占收到低保的总人数的不到 1%，可以排除残疾人因素对于“低保标准”的影响^[10]；再次，根据《2016 年最新各地城乡低保标准表》，吉林城乡低保标准为 392.5，再经调阅吉林省政府网站，吉林省城市低保标准为 350 到 435 元/月，而农村为 325/月，城市与农村差别不大，也体现出吉林省政府积极响应国家城乡一体协同发展战略^[11]。

测试的数据省份(地区)是吉林，根据吉林的六项核心指标：

表 5 吉林省六项核心指标

x_1	x_3	x_6	x_{11}	x_{12}	x_{13}
53868	19967	0.9812992	1.002884232	101.6	0.108520375

我们在多次尝试后，得出当中间隐藏层神经元数目为 5 时，模型输出计算结果相对稳定。隐藏层神经元个数为其他值时，运行结果特别不稳定，每一次的运行结果相差很大。

表 6 隐藏层神经元个数对应的结果输出

隐含层神经元的个数	10	6	5	4	3
运行结果	602.8556	576.4432	392.4423	560.0213	577.9096

输入到 RBF 模型后预测出吉林省“低保标准”平均为 400.0，与国家标准的 392.5 相差不大，模型的稳定性与可靠性得到了验证。

同时，为了再一次验证模型的可靠性，我们也对上海的数据进行测试，根据上海市的六项核心数据：

表 7 上海市六项核心指标

x_1	x_3	x_6	x_{11}	x_{12}	x_{13}
116562	54305.3	1.018410852	1.0232611203	103.2	0.1296314642

输入到 RBF 模型后预测出上海市“低保标准”平均为 765.8594，与国家标准的 790 相差不大，模型的稳定性与可靠性得到了验证。

6.3 问题 3 的求解——利用综合评价分析相关性的模型

6.3.1 模型分析与准备

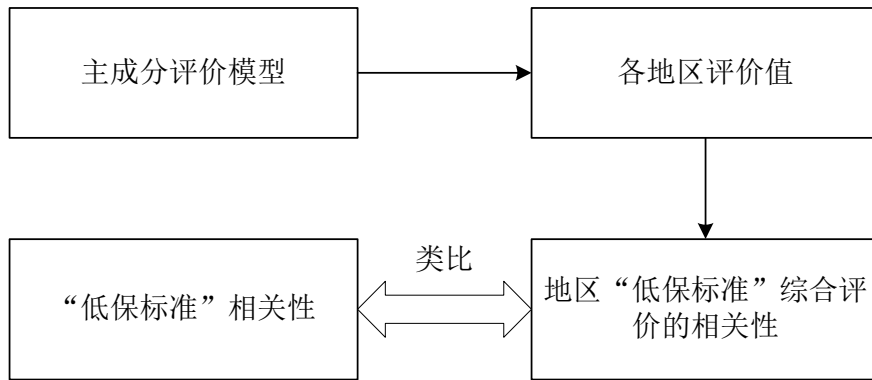


Fig. 6 模型应用结构图

6.3.2 模型建立

分析各地区“低保标准”相关性可以用关于“低保标准”的评价相关性替代，利用主成分分析的数学方法得出各地区关于“低保标准”的综合得分 Z 。

$$Z = \sum_{j=1}^p b_j y_j$$

其中： b_j 为第 j 个主成分的信息贡献率，根据综合得分进行评价，再根据评价进行相关性分类。

前面 5.1 中采用了主成分分析法求解出 6 个主成分 $y_1, y_2, y_3, y_4, y_5, y_6$ ，以表格 2 中列出的 6 个主成分的贡献率为权重，构建主成分综合评价模型，即

$$Z=0.3953y_1 + 0.1719y_2 + 0.1172y_3 + 0.0897y_4 + 0.0729y_5 + 0.04963y_6$$

把各地区的 6 个主成分值带入上式中，可以求得各地区“低保标准”的综合评价价值。下面利用 Matlab 进行计算排列，表格 8 给出了排序结果。

表 8 各地区“低保标准”排名和综合评价结果

名次	1	2	3	4	5	6	7	8
地区	上海	北京	天津	江苏	浙江	山东	辽宁	福建
综合评价价值	2.8407	1.8491	1.832	1.3567	0.7874	0.4993	0.2872	0.1951
名次	9	10	11	12	13	14	15	16
地区	内蒙古	湖北	重庆	广东	吉林	四川	黑龙江	湖南
综合评价价值	0.1394	0.067	0.0623	0.032	0.0108	0.0138	0.0728	0.1046
名次	17	18	19	20	21	22	23	24
地区	安徽	海南	河北	陕西	河南	宁夏	江西	青海
综合评价价值	0.1468	0.2153	0.2153	0.3002	0.3323	0.5354	0.5848	0.5882
名次	25	26	27	28	29	30	31	
地区	山西	广西	贵州	新疆	云南	甘肃	西藏	
综合评价价值	0.6219	0.6561	0.7377	0.8135	0.8673	1.1003	1.9902	

6.3.3 模型结果分析

从表格数据分析到地区“低保”标准存在差异，为了更直观看出差异地区，根据综合评价得分 Z 的数据用 excel 绘出散点图。

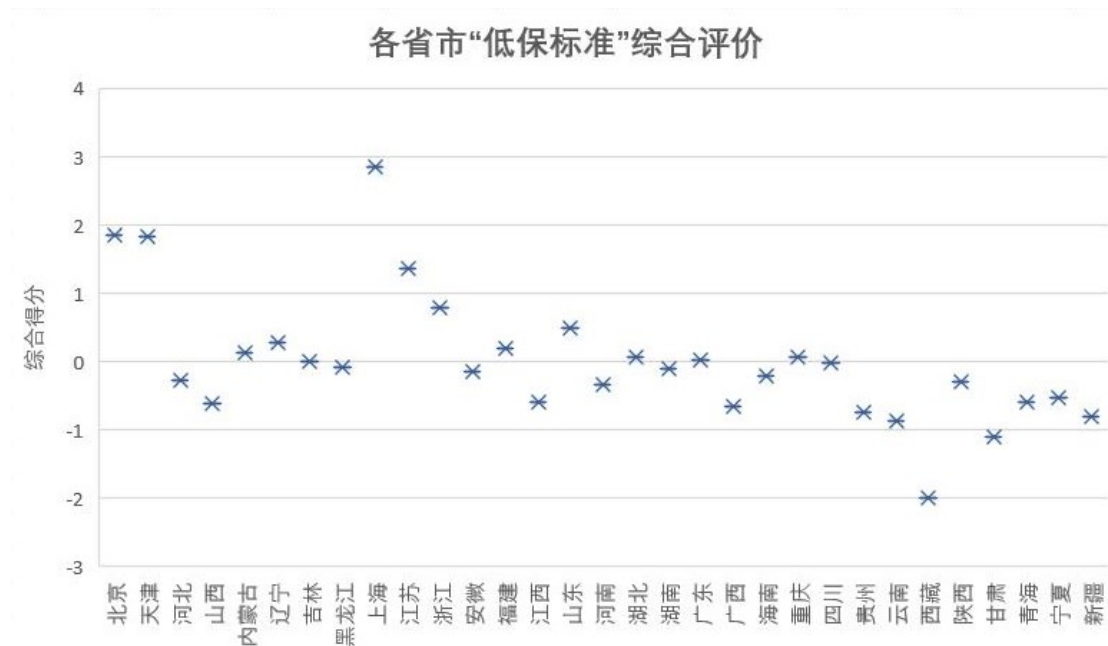


Fig. 7 各省市“低保标准”综合评价图

根据散点图我们很容易能看出“低保标准”相关性较大的省市，我们经过多次数据统计分析决定将阈值设置为 0.9999 和 -0.9999，评价结果在同一区间内的省市“低保标准”相关性高。我们根据阈值按照相关性将各省市分为三类。上海、北京、天津、江苏为 I 类，它们的“低保标准”更高，主要表现在人均国内生产总值高等方面；甘肃、西藏为 III 类，它们的“低保标准”较低；剩余的其他地区为 II 类，“低保标准”处于中等水平。

表 9 分类后的城市

类别	I 类	II 类	III 类
省市	上海、北京、天津、江苏	其余省市	甘肃、西藏

结论: 以得分评价价值高的前两位上海和北京为例，这两个城市在“低保标准”上相关性强，通过我们的对这两个大城市的了解，两个地区很多方面相似，同为中国的一线城市，也面临着大城市带来的共同问题。通过该模型给出的结果经过多方验证具有可靠性。

6.4 问题 4 的求解——同类高精度预测模型

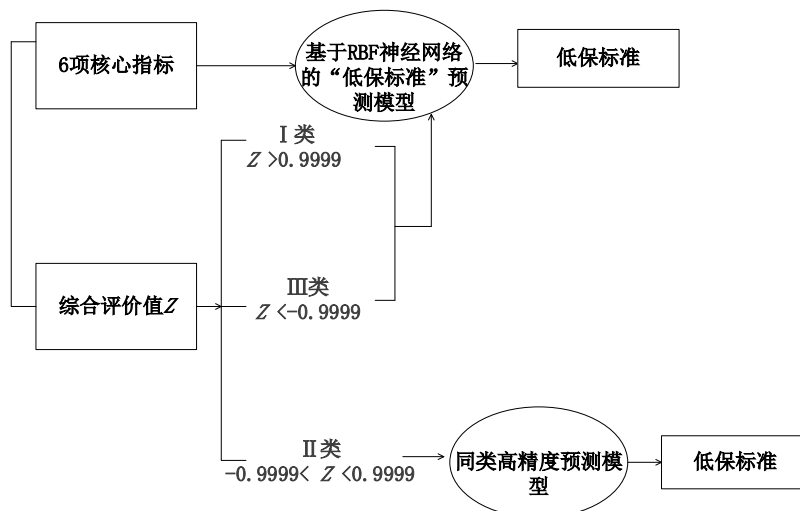


Fig. 8 思路分析图

6.4.1 模型分析与准备

我们在问题二中建立的是拟合回归全国 31 个省市的具有普遍性和灵活性的基于 RBF 神经网络的“低保标准”预测模型，但是这个模型中的部分省市由于地理位置、城市发展优先或延迟、国家政策等一系列因素，出现很多极端值：例如 I 类北京、上海、天津、江苏为第一梯队，“低保标准”遥遥领先高于全国其他省市；例如 III 类甘肃、西藏作为压低梯队，“低保标准”明显远远低于全国其他省市。

在近 6 大省市极其特殊的情况下，其 6 项指标较为“偏激”，会对整个神经网络的模型造成一定的影响，基于此，在第三问题的基础上，我们按照综合标准 Z 的取值选取阈值分类为三大类，中间一类为剩余的其余地区，共有 25 座省市，且这一大类的 25 座省市的“低保标准”没有较大或者较小波动变化，恰好适合我们再次单独利用 25 座省市的 6 项核心指标建立 RBF 神经网络模型，以期望达到最好的效果。

6.4.2 模型建立

我们所建立的同类高精度预测模型囊括：

第一步主成分分析，得出 6 项核心指标与综合评价值 Z，对 Z 值进行分类；

第二步，如果 Z 值属于 I 类或者 III 类的阵营，就将预测的省市 6 项基本指标输入到我们在第二问题中训练的基于 RBF 神经网络的“低保标准”预测模型，进行对极端值的预测；如果 Z 值属于 II 类剩余的其他类型，就将该省市输入到我们新训练的只针对绝大部分省市的 RBF 神经网络的“低保标准”预测模型；

第三步，输出结果“低保标准”，并且得出结论

我们在挑选出 25 组 II 类省市的数据之后，对模型进行了训练与预测。

6.4.3 模型适应性验证

我们的模型仍然选取 6 个（6 项核心指标）自变量为输入，1 个因变量（“低保标准”）为输出，中间隐藏层的神经元数量选取 4 或 5，在训练完模型之后，我们发现，无论中间隐藏层的神经元数目选取多少（根据经验一般选取 3~10），模型输出值都趋于稳定，这达到了我们预期效果。选取吉林与西藏作为我们的研究对象。

在将吉林的 13 项对“低保标准”造成影响的指标进行主成分分析后，得出吉林的 6 项核心指标与综合评价值 Z：

表 10 吉林省 6 项核心指标与综合评价值

x_1	x_3	x_6	x_{11}	x_{12}	x_{13}	Z
53868	19967	0.9812992	1.002884232	101.6	0.108520375	0.0108

由于吉林省得综合评价值 Z 介于 -0.9999 与 0.9999 之间，所以将其 6 项核心指标输入到我们新建立的 RBF 神经网络的“低保标准”预测模型，在进行测试之后，得出吉林的“低保标准”为 390.9854，与吉林省标准 392.5 相差不大。

在将西藏的 13 项对“低保标准”造成影响的指标进行主成分分析后，得出西藏的 6 项核心指标与综合评价值 Z：

表 11 西藏 6 项核心指标与综合评价值

x_1	x_3	x_6	x_{11}	x_{12}	x_{13}	Z
40564	18354.7	0.9980276	0.9718639	101.4	0.0723035	-1.9902

由于西藏得综合评价值 Z 小于 -0.9999 ，所以将其 6 项核心指标输入到我们在问题二建立的 RBF 神经网络的“低保标准”预测模型，在进行测试之后，得出西藏的“低保标准”为 380.0549，与西藏标准 380 相差不大。

结论：在对吉林与西藏两个省份进行分析之后，我们得出了非常精准的“低保标准”，证明了我们建立的模型的精准度与可靠性。

7. 模型的优缺点

7.1 模型的优点

1) 问题一模型建立过程中，为了让模型更加贴合实际情况，我们考虑了众多影响指标，计算的结果更符合实际。

2) 问题一模型用少数重要指标替代了大多数指标。

3) 解决问题三是建立在问题一模型的基础上，简化了问题的复杂度。

7.2 模型的缺点

1) 模型选取影响指标时忽略掉了某些特殊指标。

2) 问题三对问题一模型依赖度高，受问题一结果影响。

8. 模型的改进

针对实际问题，对模型进行下一改进：

(1) 在低保标准的讨论过程中，考虑到城乡一体化战略的大力实施，我们只分析某地区城市与农村低保相同，没有考虑极少数偏远地区地方城市与农村低保的差异。

在对问题一的解答中，我们在考察研究的基础上总结、筛选、整理出 13 项与“低保标准”有关的指标，而现实生活中，对“低保标准”有影响的因素指标众多，涉及到国家政策、自然灾害等突发情况的救助、按照季节变化的动态影响等多方面因素。故在模型的改进中，如果能够综合考虑各方面因素的影响因子，既包括静态考量，也包括动态特性影响，则可使得主成分分析模型挑选出的贡献率更大的指标更合理。

在对问题二的解答中，我们的数据集只选用 2016 年全国 31 个省市的 6 项核心指标，如果可以增加数据集，既可以增加不同年份的数据集，也可以增加市级、县级、乡级以至区级的数据，都可以使模型更加强健，使结果更加合理。

在对问题二的解答中，适当增加神经网络的训练次数与输出结果次数，去掉极端值，取平均值或中位数，可使计算机模拟的结果更加合理。

在对问题三的解答中，我们只通过问题一主成分分析得出的综合评价值 Z 进行了分类与分析，如果能够针对问题三建立一种较为合理的基于多变量评价指标的评价模型，给出评价结果，并且依据结果给出详细的优化模型方案，那样会更加合理。

9. 模型的推广

本模型是由几个子模型组成：1. 基于主成分分析“低保标准”指标挑选模型 2. 基于 RBF 神经网络的“低保标准”预测模型 3. 利用综合评价分析相关性的模型。是一套比较齐全的解决成分分析与预测的模型集。

模型所挑选的指标对全国各大省市均为“低保标准”的核心指标，具有较强的普遍性和灵活性；

1. 利用神经网络可有效解决非线性多维度预测问题；
2. 解决“低保标准”的发放金额、模型预测、优化、指标计算挑选等问题；
3. 模型的评价指标给出，以及评价方法可以推广到国家的其方面的指标的评价上；
4. 模型计算机模拟可以推广到成分分析、分类、预测中，如：国家运动员补贴发放、政府公车补贴等。

10. 参考文献

- [1] 陈仲常, 龚锐. 不同地区低保标准比较的参照基准研究[J]. 市场与人口分析, 2005, 11(6):14~17
- [2] 罗静, 沙治慧. 城乡低保待遇水平与物价联动标准测算方法改进[J]. *Social Security Studies*, 2018, 11(1):80~86
- [3] 司守奎, 孙玺菁. 数学建模算法与应用[M]. 北京:国防工业出版社, 2017:213~216
- [4] 褚福灵. 我国城乡低保标准现状评估研究[J]. 北京劳动保障职业学院学报, 2016, 10(5):8~12
- [5] Bing Wang, Yao-hua Meng, Xiao-hong Yu. Radial basis function process neural network training based on generalized frechet distance and GA-SA hybrid strategy[J]. *Computer Science & Engineering: An International Journal (CSEIJ)*, 2013, 12(3):1~9
- [6] Jifu Nong. The Design of RBF Neural Networks and experimentation for solving overfitting problem[C]. In: 2011 International Conference on Electronics and Optoelectronics (ICEOE 2011), 2011, 75~78
- [7] 汪泓, 张伯生, 郁健. 上海市最低工资标准测算模型的研究[J]. 上海工程技术大学学报, 2001, 08(15):8~15
- [8] 张立明. 人工神经网络的模型及其应用. 上海:复旦大学出版社, 1993. 44~47
- [9] 张时飞, 唐钧. 论中国城市居民的“基本生活需要”. 河北学刊, 2007(1):6~10
- [10] 汪三贵. 中国的农村贫困与社会保障. 中国社会福利, 2008, 09(7):15~21
- [11] 陈传波. 城乡居民的最低生活需求与低保标准[J]. 农业经济问题, 2011, 8(7):38~44

11. 附录

附录一：主成分分析数据集(2016年)

地区	人均生产总值/元	居民消费水平指数	人均可支配收入	食品指数	衣着指数	居住指数	生活用品及服务指数
北京	118198	48883	52530.4	1.01873767	0.988166	1.022682	0.978303748
天津	115053	36257	34074.5	1.00685602	0.980411	1.014691	0.973555338
河北	43062	14328	19725.4	1.01477833	1.002956	0.992118	0.990147783
山西	35532	15065	19048.9	1.02176063	0.999011	0.988131	0.989119683
内蒙古	72064	22293	24126.6	1.01383399	1.001976	0.988142	0.989130435
辽宁	50791	23670	26039.7	1.01377953	0.998031	0.989173	0.991141732
吉林	53868	13786	19967	1.01968504	1.001969	0.981299	0.989173228
黑龙江	40432	17393	19838.5	1.01576355	0.995074	0.985222	0.989162562
上海	116562	49617	54305.3	1.01550388	0.976744	1.018411	0.980620155
江苏	96887	35875	32070.1	1.02346041	0.995112	0.989247	0.99315738
浙江	84916	30743	38529	1.03140334	0.996075	0.991168	0.983316977
安徽	39561	15466	19998.1	1.02750491	0.990177	0.993124	0.984282908
福建	74707	23355	27607.9	1.02949853	0.986234	0.990167	0.981317601
江西	40400	16040	20109.6	1.03529412	0.989216	0.990196	0.981372549
山东	68733	25860	24685.3	1.0215475	0.996082	0.988247	0.987267385
河南	42575	16043	18443.1	1.02158979	0.988224	1.002944	0.983316977
湖北	55665	19391	21786.6	1.03131115	1.000978	1.005871	0.982387476
湖南	46382	17490	21114.8	1.03336605	0.996075	0.993131	0.981354269
广东	74016	28495	30295.8	1.03519062	1.00391	0.994135	0.979472141
广西	38027	15013	18305.1	1.0265748	0.997047	0.987205	0.983267717
海南	44347	18431	20653.4	1.03210117	0.952335	0.998054	0.980544747
重庆	58502	21032	22034.1	1.02848723	1.005894	0.993124	0.988212181
四川	40003	16013	18808.3	1.03042198	0.987242	0.993131	0.984298332
贵州	33246	14666	15121.1	1.03057199	0.982249	0.994083	0.985207101
云南	31093	14534	16719.9	1.02758621	0.987192	0.997044	0.985221675
西藏	35184	9743	13639.2	1.01853659	1.006829	0.983415	0.990243902
陕西	51015	16657	18873.7	1.02467917	0.998026	0.996051	0.982230997
甘肃	27643	13086	14670.3	1.0266535	1.000987	0.995064	0.991115499
青海	43531	16751	17301.8	1.00589391	0.994106	1.033399	0.986247544
宁夏	47194	18570	18832.3	1.01674877	1.00197	0.989163	0.98817734
新疆	40564	15247	18354.7	1.00098619	0.999014	0.998028	0.992110454

地区	交通和通信指数	教育文化与娱乐指数	医疗保健指数	家庭户数指数	物价指数	人口老龄化指数
北京	0.952662722	0.969428008	1.01183432	1.014673572	101.4	0.117416722
天津	0.962781587	0.985308521	1.065621939	1.009732207	102.1	0.113598038
河北	0.968472906	0.998029557	1.028571429	0.980792473	101.5	0.109083665
山西	0.972304649	1.001978239	1.012858556	0.994291564	101.1	0.086962148
内蒙古	0.977272727	0.995059289	1.031620553	1.009034418	101.2	0.094388721
辽宁	0.982283465	1.011811024	1.008858268	1.009329287	101.6	0.132240646
吉林	0.971456693	0.99015748	1.044291339	1.002884232	101.6	0.108520375
黑龙江	0.985221675	1.001970443	1.021674877	1.006743455	101.5	0.119376294
上海	0.939922481	0.995155039	1.05620155	1.02326112	103.2	0.129631464
江苏	0.965786901	0.986314761	1.066471163	0.986459347	102.3	0.135048807
浙江	0.968596663	1.007850834	0.994111874	1.013943395	101.9	0.116354551
安徽	0.958742633	1.004911591	1.017681729	0.975048929	101.8	0.114664976
福建	0.977384464	0.995083579	1.01179941	0.990406402	101.7	0.099494981
江西	0.968627451	0.995098039	1.006862745	0.950618203	102	0.095966404
山东	0.975514202	0.998041136	1.027424094	1.0051857	102.1	0.116517301
河南	0.964671246	1.004906771	1.008832188	0.96964658	101.9	0.100761168
湖北	0.951076321	1	0.997064579	0.991261108	102.2	0.115745991
湖南	0.965652601	0.989205103	1.011776251	0.981090502	101.9	0.118897226
广东	0.96285435	0.991202346	1.004887586	0.974315971	102.3	0.076932264
广西	0.972440945	1	1.020669291	0.947518843	101.6	0.096934386
海南	0.958171206	1.001945525	1.015564202	0.936757426	102.8	0.082488958
重庆	0.988212181	0.97740668	1	1.010077749	101.8	0.139671362
四川	0.967615309	1.005888126	0.997055937	0.995809415	101.9	0.136962437
贵州	0.973372781	0.999013807	1.000986193	0.970676734	101.4	0.096239345
云南	0.978325123	0.992118227	1.008866995	0.962898564	101.5	0.083754764
西藏	0.970731707	0.986341463	0.996097561	0.868488889	102.5	0.049838652
陕西	0.970384995	0.987166831	1.010858835	0.987134595	101.3	0.107421753
甘肃	0.977295163	0.987166831	0.995064166	0.962878955	101.3	0.099544627
青海	0.955795678	0.988212181	1.007858546	0.966125085	101.8	0.072187688
宁夏	0.971428571	1.006896552	1.013793103	0.984228604	101.5	0.077832686
新疆	0.979289941	1.001972387	1.012820513	0.9718639	101.4	0.072303496

附录二：《2016 年最新各地城乡低保标准表》

地区	城市低保标准
上海	790
北京	710
天津	705
南京	700
杭州	660
拉萨	640
广州	600
武汉	580
郑州	520
哈尔滨	510
合肥	510
石家庄	500
呼和浩特	515-565
沈阳	505-580
济南	500-550
西安	480-510
昆明	475-530
太原	453-505
南昌	450-480
长沙	450
海口	450
贵阳	425-530
成都	400-500
兰州	387-515
银川	380
乌鲁木齐	380
西宁	373

重庆	365-385
长春	350-435
南宁	250
福州	最低工资标准 36%-42%

附录三：主成分分析代码

```

%=====
clc, clear

load data.txt          %把原始数据保存在纯文本文件 data.txt 中
data=zscore(data);    %数据标准化
r=corrcoef(data);     %计算相关系数矩阵

%下面利用相关系数矩阵进行主成分分析，vec1 的列为 r 的特征向量，即主成分的
%系数, lamda 为 r 的特征值，rate 为各个主成分的贡献率

[vec1, lamda, rate]=pcacov(r)

%构造与 vec1 同维数的元素为±1 的矩阵
f= repmat(sign(sum(vec1)), size(vec1, 1), 1);

%修改特征向量的正负号，使得每个特征向量的分量和为正
vec2=vec1.*f

num=4;                %num 为选取的主成分的个数
df=data*vec2(:, 1:num); %计算各个主成分的分值
tf=df*rate(1:num)/100; %计算综合得分
[stf, ind]=sort(tf, 'descend'); %把得分按照从高到低的次序排列
stf=stf', ind=ind'

%=====

```

附录四：RBF 神经网络模型代码

```

%=====
clc, clear

```

```

a=load('s2.txt');           %把表中数据保存到纯文本文件
a=a';                       %矩阵转置
P=a([1:6],[1:end-1]);[PN,PS1]=mapminmax(P); %自变量数据规格化到
[-1,1]
T=a(7,[1:end-1]);[TN,PS2]=mapminmax(T); %因变量数据规格化到[-1,1]
net1=newrb(PN,TN)           %训练 RBF 网络
x=a([1:6],end);xn=mapminmax('apply',x,PS1); %预测样本点自变量规格化
%求预测值并把数据还原
yn1=sim(net1,xn); y1=mapminmax('reverse',yn1,PS2)
delta1=abs(a(7,31)-y1)/a(7,31) %计算 RBF 网络预测的相对误差
%初始化 BP 网络，隐含层的神经元取为 4 个（多次试验）
net2=feedforwardnet(4);
net2=train(net2,PN,TN); %训练 BP 网络
%求预测值，并把数据还原
yn2= net2(xn);y2=mapminmax('reverse',yn2,PS2)
%=====

```

附录五：RBF 神经网络训练&测试集

地区	人均生产总值/元	人均可支配收入	居住指数	家庭户数指数	物价指数	人口老龄化指数	低保标准
北京	118198	52530.4	1.0226 824	1.0146735 72	101.4	0.11741672 2	710
天津	115053	34074.5	1.0146 915	1.0097322 07	102.1	0.11359803 8	705
河北	43062	19725.4	0.9921 182	0.9807924 73	101.5	0.10908366 5	500
山西	35532	19048.9	0.9881 306	0.9942915 64	101.1	0.08696214 8	479
内蒙古	72064	24126.6	0.9881 423	1.0090344 18	101.2	0.09438872 1	540
辽宁	50791	26039.7	0.9891 732	1.0093292 87	101.6	0.13224064 6	542.5
吉林	53868	19967	0.9812 992	1.0028842 32	101.6	0.10852037 5	392.5
黑龙江	40432	19838.5	0.9852 217	1.0067434 55	101.5	0.11937629 4	510

上海	116562	54305.3	1.0184 109	1.0232611 2	103.2	0.12963146 4	790
江苏	96887	32070.1	0.9892 473	0.9864593 47	102.3	0.13504880 7	700
浙江	84916	38529	0.9911 678	1.0139433 95	101.9	0.11635455 1	660
安徽	39561	19998.1	0.9931 238	0.9750489 29	101.8	0.11466497 6	510
福建	74707	27607.9	0.9901 672	0.9904064 02	101.7	0.09949498 1	700
江西	40400	20109.6	0.9901 961	0.9506182 03	102	0.09596640 4	465
山东	68733	24685.3	0.9882 468	1.0051857	102.1	0.11651730 1	525
河南	42575	18443.1	1.0029 441	0.9696465 8	101.9	0.10076116 8	520
湖北	55665	21786.6	1.0058 708	0.9912611 08	102.2	0.11574599 1	580
湖南	46382	21114.8	0.9931 305	0.9810905 02	101.9	0.11889722 6	450
广东	74016	30295.8	0.9941 349	0.9743159 71	102.3	0.07693226 4	600
广西	38027	18305.1	0.9872 047	0.9475188 43	101.6	0.09693438 6	250
海南	44347	20653.4	0.9980 545	0.9367574 26	102.8	0.08248895 8	450
重庆	58502	22034.1	0.9931 238	1.0100777 49	101.8	0.13967136 2	375
四川	40003	18808.3	0.9931 305	0.9958094 15	101.9	0.13696243 7	450
贵州	33246	15121.1	0.9940 828	0.9706767 34	101.4	0.09623934 5	477.5
云南	31093	16719.9	0.9970 443	0.9628985 64	101.5	0.08375476 4	502.5
西藏	35184	13639.2	0.9834 146	0.8684888 89	102.5	0.04983865 2	640
陕西	51015	18873.7	0.9960 513	0.9871345 95	101.3	0.10742175 3	495
甘肃	27643	14670.3	0.9950 642	0.9628789 55	101.3	0.09954462 7	451
青海	43531	17301.8	1.0333 988	0.9661250 85	101.8	0.07218768 8	373
宁夏	47194	18832.3	0.9891 626	0.9842286 04	101.5	0.07783268 6	380

新疆	40564	18354.7	0.9980 276	0.9718639	101.4	0.07230349 6	380
----	-------	---------	---------------	-----------	-------	-----------------	-----